

Review

ϕ -Divergence in Contingency Table Analysis

Maria Kateri

Institute of Statistics, RWTH Aachen University, 52062 Aachen, Germany; maria.kateri@rwth-aachen.de

Received: 6 April 2018; Accepted: 20 April 2018; Published: 27 April 2018



Abstract: The ϕ -divergence association models for two-way contingency tables is a family of models that includes the association and correlation models as special cases. We present this family of models, discussing its features and demonstrating the role of ϕ -divergence in building this family. The most parsimonious member of this family, the model of ϕ -scaled uniform local association, is considered in detail. It is implemented and representative examples are commented on.

Keywords: log-linear models; ordinal classification variables; association models; correlation models

1. Introduction

Contingency tables and their analysis are of special importance for various diverse fields, like medical sciences, psychology, education, demography and social sciences. In these fields, categorical variables and their cross-classification play a predominant role, since characteristics of interest are often categorical, nominal or most frequently ordinal. For example, diagnostic ratings, strength of opinions or preferences, educational attainments and socioeconomic characteristics are expressed in ordinal scales. The origins of contingency table analysis (CTA) lie back in 1900 with the well-known contributions by Pearson and Yule, while, for the history of CTA before 1900, we refer to the interesting paper by Stigler [1]. The interest lies mainly in identifying and describing structures of underlying association in terms of appropriate models or measures. Divergence measures have been employed in the CTA mainly for hypothesis testing (model fit) and estimation, leading to general families of test statistics and estimators. The family of ϕ -divergence test statistics contain the classical likelihood ratio and Pearson test statistics as special cases while the maximum likelihood estimators (MLEs) belong to the family of the minimum ϕ -divergence estimators (M ϕ Es). Families of ϕ -divergence based test statistics as well as M ϕ Es for various standard models in CTA have a long history. However, their consideration and discussion is out of our scope. For a detailed overview and related references, we refer to the comprehensive book of Pardo [2]. For log-linear models, see also [3].

Here, we aim at highlighting a different structural role of ϕ -divergence in contingency tables modelling, namely that of linking phenomenological different models, forming thus a family of models and providing a basis for their comparison, understanding and unified treatment. Through this approach, new insight is gained for the standard association and correlation models (see [4,5]) while further alternatives are considered. We restrict our discussion on two-dimensional contingency tables, but the models and approaches discussed are directly extendable to tables of higher dimension.

The organization of the paper is as follows. Preliminaries on log-linear models, divergence measures, association and correlation models for two-way tables are provided in Section 2. In the sequel, the general family of ϕ -divergence based association models (AMs), which includes the classical association and correlation models as special cases, is reviewed in Section 3. The most parsimonious ϕ -divergence based association model, that of ϕ -scaled uniform local association, and its role in conditional testing of independence is considered and discussed in Section 4. For this family of models, the effect of the specific ϕ -function used is illustrated by analysing representative examples in Section 5. Some final comments are provided in Section 6.

2. Preliminaries

Consider an $I \times J$ contingency table $\mathbf{n} = (n_{ij})$ with rows and columns classification variables X and Y , respectively, where n_{ij} is the observed frequency in cell (i, j) . The total sample size $n = \sum_{i,j} n_{ij}$ is fixed and the random table \mathbf{N} is multinomial distributed $\mathbf{N} \sim \mathcal{M}(n, \boldsymbol{\pi})$, with probability table $\boldsymbol{\pi} \in \Delta_{IJ}$, where Δ_{IJ} is the simplex $\Delta_{IJ} = \{\boldsymbol{\pi} = (\pi_{ij}) : \pi_{ij} > 0, \sum_{i,j} \pi_{ij} = 1\}$. Let $\mathbf{m} = E(\mathbf{N}) = n\boldsymbol{\pi}$ be the table of expected cell frequencies. Since the mapping $(n, \boldsymbol{\pi}) \mapsto \mathbf{m}$ is one-to-one on Δ_{IJ} , $\mathbf{m} \in \{\mathbf{m} = (m_{ij}) : m_{ij} > 0, \sum_{i,j} m_{ij} = n\}$ and models (hypotheses) for $\boldsymbol{\pi}$ can equivalently be expressed in terms of \mathbf{m} . Furthermore, let $\boldsymbol{\pi}_r = (\pi_{1+}, \dots, \pi_{I+})^T$ and $\boldsymbol{\pi}_c = (\pi_{+1}, \dots, \pi_{+J})^T$ be the row and column marginal probabilities vectors, respectively, and $\mathbf{p} = (p_{ij})$ the table of sample proportions with $p_{ij} = n_{ij}/n$.

The classical independence hypothesis for the classification variables X and Y ($\boldsymbol{\pi} = \boldsymbol{\pi}_r \boldsymbol{\pi}_c^T = \boldsymbol{\pi}^I$) corresponds to the log-linear model of independence (I), defined in terms of expected cell frequencies as

$$\log(m_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y, \quad i = 1, \dots, I, j = 1, \dots, J, \quad (1)$$

where λ_i^X and λ_j^Y are the i -th row and j -th column main effects, respectively, while λ is the intercept. If the independence model is rejected, the interaction between X and Y is significant and the only alternative in the standard log-linear models set-up is the saturated model

$$\log(m_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \lambda_{ij}^{XY}, \quad i = 1, \dots, I, j = 1, \dots, J, \quad (2)$$

which imposes no structure on $\boldsymbol{\pi}$. Identifiability constraints need to be imposed on the main effect and interaction parameters of these models, like $\lambda_1^X = \lambda_1^Y = 0$ and $\lambda_{1j}^{XY} = \lambda_{i1}^{XY} = 0$, for all i, j .

An important generalized measure for measuring the divergence between two probability distributions is the ϕ -divergence. Let $\boldsymbol{\pi} = (\pi_{ij})$, $\mathbf{q} = (q_{ij}) \in \Delta_{IJ}$ be two discrete finite bivariate probability distributions. Then, the ϕ -divergence between \mathbf{q} and $\boldsymbol{\pi}$ (or Csiszar's measure of information in \mathbf{q} about $\boldsymbol{\pi}$), is given by

$$I_\phi^C(\mathbf{q}, \boldsymbol{\pi}) = \sum_{i,j} \pi_{ij} \phi(q_{ij}/\pi_{ij}), \quad (3)$$

where ϕ is a real-valued strictly convex function on $[0, \infty)$ with $\phi(1) = \phi'(1) = 0$, $0\phi(0/0) = 0$, $0\phi(y/0) = \lim_{x \rightarrow \infty} \phi(x)/x$ (cf. [2]). Setting $\phi(x) = x \log x$, (3) is reduced to the Kullback–Leibler (KL) divergence

$$I^{\text{KL}}(\mathbf{q}, \boldsymbol{\pi}) = \sum_{i,j} q_{ij} \log(q_{ij}/\pi_{ij}), \quad (4)$$

while, for $\phi(x) = (1-x)^2$, Pearson's divergence is derived. If $\phi(x) = \frac{x^{\lambda+1}-x}{\lambda(\lambda+1)}$, (3) becomes the power divergence measure of Cressie and Read [6]

$$I_\lambda^{\text{CR}}(\mathbf{q}, \boldsymbol{\pi}) = \frac{1}{\lambda(\lambda+1)} \sum_{i=1}^K q_i \left[\left(\frac{q_i}{\pi_i} \right)^\lambda - 1 \right], \quad -\infty < \lambda < \infty, \quad \lambda \neq -1, 0. \quad (5)$$

For $\lambda \rightarrow -1$ and $\lambda \rightarrow -0$, (5) converges to the $I^{\text{KL}}(\boldsymbol{\pi}, \mathbf{q})$ and $I^{\text{KL}}(\mathbf{q}, \boldsymbol{\pi})$, respectively, while $\lambda = 1$ corresponds to Pearson's divergence.

The goodness of fit (GOF) of model (1) is usually tested by the likelihood ratio test statistic $G^2 = 2nI^{\text{KL}}(\mathbf{p}, \hat{\boldsymbol{\pi}})$ or Pearson's $X^2 = 2nI_1^{\text{CR}}(\mathbf{p}, \hat{\boldsymbol{\pi}})$, where $\hat{\boldsymbol{\pi}}$ is the MLE of $\boldsymbol{\pi}$ under (1). Both test statistics are under (1) asymptotically $\chi_{(I-1)(J-1)}^2$ distributed. Alternatively, ϕ -divergence test statistics can be used (see [3]).

2.1. Association Models

In case of ordinal classification variables, the association models (AMs) impose a special structure on the underlying association and thus provide non-saturated models of dependence. AMs are based on scores $\mu = (\mu_1, \dots, \mu_I)$ and $\nu = (\nu_1, \dots, \nu_J)$ assigned to the rows and columns of the ordinal classification variables, respectively. They are defined by the expression

$$\log(m_{ij}) = \lambda + \lambda_i^X + \lambda_j^Y + \zeta \mu_i \nu_j, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad (6)$$

where the row and column scores are standardized subject to weights $w_1 = (w_{11}, \dots, w_{1I})^T$ and $w_2 = (w_{21}, \dots, w_{2J})^T$, respectively, i.e., it holds

$$\sum_i w_{1i} \mu_i = \sum_j w_{2j} \nu_j = 0 \quad \text{and} \quad \sum_i w_{1i} \mu_i^2 = \sum_j w_{2j} \nu_j^2 = 1. \quad (7)$$

Usually, the uniform ($w_1 = \mathbf{1}_I$, $w_2 = \mathbf{1}_J$, where $\mathbf{1}_k$ is a $k \times 1$ vector of 1s) or the marginal weights ($w_1 = \pi_r$, $w_2 = \pi_c$) are used.

If μ and ν are both known and ordered, then (6) has just one parameter more than independence, parameter ζ , and is known as the *Linear-by-Linear* (LL) AM. In case the vector μ is unknown, (6) is the *Row effect* (R) AM, while the *Column effect* (C) AM is defined analogously. Finally, when the row and the column scores are all unknown parameters to be estimated, (6) is the *multiplicative Row-Column* (RC) AM. Scores that are unknown need not necessarily be ordered. Note that models LL, R and C are log-linear while the RC is not. The degrees of freedom (df) of these AMs equal $df(\text{LL}) = (I-1)(J-1)-1$, $df(\text{R}) = (I-1)(J-2)$, $df(\text{C}) = (I-2)(J-1)$ and $df(\text{RC}) = (I-2)(J-2)$. The special LL model for which the row and column scores are equidistant for successive categories is known as the *Uniform* (U) AM.

In case the RC model is not of adequate fit, multiplicative row-column AMs of higher order can be considered. Such a model of M -th order is defined as

$$\log m_{ij} = \lambda + \lambda_i^X + \lambda_j^Y + \sum_{m=1}^M \zeta_m \mu_{im} \nu_{jm}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad (8)$$

with $1 \leq M \leq M^* = \min(I, J) - 1$, and denoted by $\text{RC}(M)$. Model $\text{RC}(M^*)$ is an equivalent expression of the saturated model (2). The sum $\sum_{m=1}^M \zeta_m \mu_{im} \nu_{jm}$ in (8) corresponds to the generalized singular value decomposition of the matrix of interaction parameters of model (2), $\Lambda = (\lambda_{ij}^{XY})_{I \times J}$, and M is the rank of Λ . The ζ_m s are the associated eigenvalues, satisfying thus $\zeta_1 \geq \dots \geq \zeta_M > 0$. Vectors $\mu_m = (\mu_{1m}, \dots, \mu_{Im})$ and $\nu_m = (\nu_{1m}, \dots, \nu_{Jm})$ are the corresponding row and column eigenvectors, $m = 1, \dots, M$, which are orthonormalized with respect to the weights w_1 and w_2 , i.e., the following constraints are satisfied:

$$\begin{aligned} \sum_i w_{1i} \mu_{im} &= \sum_j w_{2j} \nu_{jm} = 0, \quad m = 1, \dots, M, \\ \sum_i w_{1i} \mu_{im} \mu_{i\ell} &= \sum_j w_{2j} \nu_{jm} \nu_{j\ell} = \delta_{m\ell}, \quad m, \ell = 1, \dots, M, \end{aligned} \quad (9)$$

where $\delta_{m\ell}$ is Kronecker's delta. It can easily be verified that $df(\text{RC}(M)) = (I-M-1)(J-M-1)$. AMs have been mainly developed by Goodman (see [4,5] and references therein) and are thus often referred to as Goodman's AMs. For a detailed presentation of the association models, their inference, properties, interpretation, the role of the weights used and associated literature, we refer to the book of Kateri [7] (Chapter 6).

For ease in understanding but also for interpretation purposes, it is convenient to think in terms of the local associations of the table and define thus AMs through the local odds ratios (LORs)

$$\theta_{ij} = \frac{\pi_{ij}\pi_{i+1,j+1}}{\pi_{i+1,j}\pi_{i,j+1}}, \quad i = 1, \dots, I-1, j = 1, \dots, J-1. \quad (10)$$

Recall that the $(I-1) \times (J-1)$ table of LORs $\theta = (\theta_{ij})$, jointly with the marginal probabilities vectors π_r and π_c , specify uniquely the corresponding $I \times J$ probability table π . Thus, given π_r and π_c , a model on π can equivalently be expressed in terms of θ . Hence, model (8) is alternatively defined as

$$\log \theta_{ij} = \sum_{m=1}^M \zeta_m (\mu_{im} - \mu_{i+1,m})(v_{jm} - v_{j+1,m}), \quad (11)$$

for $i = 1, \dots, I-1, j = 1, \dots, J-1$. In this set-up, the differences of successive row and column scores are constant for the U model, equal to

$$\mu_i - \mu_{i+1} = \Delta_1 \quad \text{and} \quad v_j - v_{j+1} = \Delta_2, \quad i = 1, \dots, I-1, j = 1, \dots, J-1, \quad (12)$$

since scores for successive categories are equidistant. Hence, the U model is equivalently defined as

$$\log \theta_{ij} = \zeta(\mu_i - \mu_{i+1})(v_j - v_{j+1}) = \zeta \Delta_1 \Delta_2 = \log \theta = \theta^{(0)} \quad (13)$$

and is the model under which all local odds ratios are equal across the table, justifying its ‘uniform association’ characterization.

2.2. Correlation Models

A popular, mainly descriptive method for exploring the pattern of association in contingency tables is correspondence analysis (CA). The detailed discussion of CA is beyond our scope. For this, we refer to the book of Greenacre [8]. Correspondence analysis is a reparameterized version of the canonical correlation model of order M

$$p_{ij} = p_{i+}p_{+j} \left(1 + \sum_{m=1}^M \rho_m x_{im} y_{jm} \right), \quad i = 1, \dots, I, j = 1, \dots, J, \quad (14)$$

with $1 \leq M \leq M^*$. The row and column scores ($\mathbf{x}_m = (x_{1m}, \dots, x_{Im})$ and $\mathbf{y}_m = (y_{1m}, \dots, y_{Jm})$, $m = 1, \dots, M$) satisfy constraints (9) subject to the marginal weights. Usually, it is assumed $M = 2$ and the row and column scores (coordinates) are graphically displayed as points in two-dimensional plots. The similarities between (8), expressed in terms of π in its multiplicative form, and (14) are obvious. Thus, motivated by the inferential approaches for AMs, MLEs have been considered for model (14), leading to the row-column correlation model of order M , while, for $M = 1$, special correlation models of U, R or C type have also been discussed by Goodman [4,5].

3. ϕ -Divergence Based Association Models

AMs and correlation models were developed competitively and opposed to each other (cf. [5]), until Gilula et al. [9] linked them in an inspiring manner under an information theoretical approach. They proved that, under certain (common) conditions, both of them are the closest model to independence. Their difference lies on the divergence used for measuring their closeness to independence. AMs are the closest in terms of the KL divergence and correlation models in terms of the Pearson’s divergence. This result motivated subsequent research and led to the definition of general classes of dependence models by substituting the KL and Pearson divergences through generalized families of divergences.

Under the conditions of [9], namely for given marginal distributions (π_r and π_c), given scores (μ and ν) and given their correlation $\rho = \text{corr}(\mu, \nu)$, Kateri and Papaioannou [10] derived that the joint distribution π that is closest to independence in terms of the ϕ -divergence is of the form

$$\pi_{ij} = \pi_{i+}\pi_{+j}F^{-1}(\alpha_i + \beta_j + \zeta\mu_i\nu_j), \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad (15)$$

where F^{-1} is the inverse function of $F(x) = \phi'(x)$. The scores μ and ν satisfy the constraints (7) with marginal weights. Under these constraints, it can easily be verified that $\rho = \text{corr}(\mu, \nu) = \sum_{i,j} \mu_i\nu_j\pi_{ij}$. Additionally, the identifiability constraints

$$\sum_i \pi_{i+}\alpha_i = \sum_j \pi_{+j}\beta_j = 0 \quad (16)$$

are imposed on parameters $\alpha = (\alpha_1, \dots, \alpha_I)$ and $\beta = (\beta_1, \dots, \beta_J)$. The parameter ζ is measuring association. It can be verified that (15), due to (7) with marginal weights and (16), leads to

$$\zeta(\pi, \mu, \nu) = \sum_{i,j} \pi_{i+}\pi_{+j}\mu_i\nu_j F\left(\frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}\right) \quad (17)$$

and that $\zeta = 0$ if and only if the independence model holds ($\pi = \pi^I$). Furthermore, under model (15), the correlation ρ between the row and column scores is increasing in ζ and the ϕ -divergence measure $I_\phi^C(\pi, \pi^I)$, for given ϕ -function, is increasing in $|\zeta|$.

Model (15), with known row and column scores, is the ϕ -divergence based extension of the LL model and is denoted by LL_ϕ . If the scores are additionally equidistant for successive categories, (15) becomes the ϕ -divergence based U model, U_ϕ , while the classes of models R_ϕ , C_ϕ and RC_ϕ are defined analogously. The standard LL, R, C or RC models correspond to $\phi(x) = x \log x$. The analogue correlation models, defined by (14) for $M = 1$, are derived for $\phi(x) = (1-x)^2$, setting $\mu = \mathbf{x}_1$, $\nu = \mathbf{y}_1$ and $\zeta = \rho_1$. For the standard association and correlation models, (17) simplifies to

$$\zeta(\pi, \mu, \nu) = \sum_{i,j} \pi_{i+}\pi_{+j}\mu_i\nu_j \log(\pi_{ij}) \quad (18)$$

and

$$\zeta(\pi, \mu, \nu) = \sum_{i,j} \mu_i\nu_j\pi_{ij} = \text{corr}(\mu, \nu) = \rho, \quad (19)$$

respectively.

For the power divergence (for $\phi(x) = \frac{x^{\lambda+1}-x}{\lambda(\lambda+1)}$, $\lambda \neq -1, 0$), model (15) becomes

$$\pi_{ij} = \pi_{i+}\pi_{+j} \left[\frac{1}{\lambda+1} + \lambda(\alpha_i + \beta_j + \zeta\mu_i\nu_j) \right]^{1/\lambda}, \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad (20)$$

considered by Rom and Sarkar [11]. Expression (20) defines parametric classes of AMs, controlled by the parameter λ , which are denoted by LL_λ , U_λ , R_λ , C_λ or RC_λ , according to the assumption made for the row and column scores.

The $RC(M)$ model, $1 \leq M \leq M^*$, is analogously generalized to $RC_\phi(M)$, the class of ϕ -divergence AMs of order M , given by

$$\pi_{ij} = \pi_{i+}\pi_{+j}F^{-1}\left(\alpha_i + \beta_j + \sum_{m=1}^M \zeta_m\mu_{im}\nu_{jm}\right), \quad i = 1, \dots, I, \quad j = 1, \dots, J, \quad (21)$$

where the scores μ_m and ν_m satisfy restrictions (9) with marginal weights. The standard RC(M) model (8) is derived for $\phi(x) = x \log x$, while, for $\phi(x) = (1-x)^2$, model (21) becomes the correlation model (14) with $\mu_m = \mathbf{x}_m$, $\nu_m = \mathbf{y}_m$ and $\zeta_m = \rho_m$, $m = 1, \dots, M$.

Models (15) and (21) can alternatively be expressed as

$$\theta_{ij}^{(\phi)} = \zeta(\mu_i - \mu_{i+1})(\nu_j - \nu_{j+1}), \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1, \quad (22)$$

and

$$\theta_{ij}^{(\phi)} = \sum_{m=1}^M \zeta_m(\mu_{im} - \mu_{i+1,m})(\nu_{jm} - \nu_{j+1,m}), \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1, \quad (23)$$

respectively, where $\theta_{ij}^{(\phi)}$ is a scaled measure of local dependence, defined for $i = 1, \dots, I-1$, $j = 1, \dots, J-1$ as

$$\theta_{ij}^{(\phi)}(\boldsymbol{\pi}) = F\left(\frac{\pi_{ij}}{\pi_{i+} \pi_{+j}}\right) + F\left(\frac{\pi_{i+1,j+1}}{\pi_{i+1,+} \pi_{+,j+1}}\right) - F\left(\frac{\pi_{i+1,j}}{\pi_{i+1,+} \pi_{+j}}\right) - F\left(\frac{\pi_{i,j+1}}{\pi_{i+} \pi_{+,j+1}}\right). \quad (24)$$

For $\phi(x) = x \log x$, (24) becomes the well-known log local odds ratio $\log(\theta_{ij})$, modelled in (13) and from now on denoted as $\theta_{ij}^{(0)} = \log(\theta_{ij})$. For the power divergence, we get

$$\theta_{ij}^{(\lambda)}(\boldsymbol{\pi}) = \frac{1}{\lambda} \left[\left(\frac{\pi_{ij}}{\pi_{i+} \pi_{+j}} \right)^\lambda + \left(\frac{\pi_{i+1,j+1}}{\pi_{i+1,+} \pi_{+,j+1}} \right)^\lambda - \left(\frac{\pi_{i+1,j}}{\pi_{i+1,+} \pi_{+j}} \right)^\lambda - \left(\frac{\pi_{i,j+1}}{\pi_{i+} \pi_{+,j+1}} \right)^\lambda \right]. \quad (25)$$

Remark 1. Kateri and Papaioannou [10] studied properties of the class of ϕ -divergence based AMs. Additionally, they developed a test based on a ϕ -divergence statistic for testing the GOF of ϕ -divergence AMs and studied its efficiency. The choice of the ϕ -function in the test statistic is independent of the ϕ -function used for the model definition. Thus, it serves as a ϕ -divergence based GOF test for the traditional well-known association or correlation models.

In order to understand the role and nature of the scaled measures of local association (24), one can examine a simple 2×2 contingency table. In this case, (10) becomes the well-known odds ratio $\theta (= \theta_{11})$. Its ϕ -divergence scaled generalization, (24) for $I = J = 2$, has been explored by Esendiller and Kateri [12]. For these ϕ -scaled association measures for 2×2 tables, asymptotic tests of significance and confidence intervals (CIs) are constructed based on the fact that they are asymptotically normal distributed. An interesting feature of the family of ϕ -scaled association measures for 2×2 tables is that when the odds ratio θ cannot be finitely estimated (due to sampling zeros), some members of this family provide finite estimates, while, for a subset of them, their variance is also finite. Extensive evaluation studies verified earlier statements in the literature about the low coverage of the log odds ratio CIs when the association is very high ($\log \theta > 4$). In such cases, focusing on the power divergence scaled odds ratios $\theta^{(\lambda)}$, $\theta^{(1/3)}$ for $\lambda = 1/3$ is to be preferred, since the corresponding CI is of better coverage when approaching the borders of the parameter space and is in general less conservative than the classical log odds ratio CI [12]. Here, the role of the scale effect on measuring the local dependence will be further clarified in the examples discussed in Section 5.

Remark 2. The idea of viewing a model as a departure from a parsimonious reference model, with the property of being the closest to this reference model under certain conditions in terms of the KL divergence, can be adopted for other types of models as well, such as the quasi symmetry (QS) model and the logistic regression, lightening thus a different interpretational aspect of these models. Substitution of the KL divergence by the ϕ -divergence, leads further, in an analogous manner to AMs, to the derivation of generalized ϕ -divergence based classes of QS models [13], ordinal QS [14] and logistic regression models [15]. For example, in case of the QS, the role of the scaled measures (24) take the analogously defined scaled deviations from the model of complete symmetry.

4. Uniform Local Association

We shall focus on the case of uniform local association and the corresponding ϕ divergence scaled model U_ϕ , defined by (15) or (22) with row and column scores satisfying (12). Model U_ϕ is equivalently expressed as

$$\theta_{ij}^{(\phi)} = \zeta \Delta_1 \Delta_2 = \theta^{(\phi)}, \quad i = 1, \dots, I-1, \quad j = 1, \dots, J-1 \quad (26)$$

and forms a family of models. Compare to (13) for the U model defined in terms of the usual log odds ratio. The associated probability table under model U_ϕ is uniquely specified by the one-to-one map $(\pi_r, \pi_c, \theta^{(\phi)}) \mapsto \pi^{U_\phi}$, not given in a closed-form expression.

The MLEs of the marginal probabilities are the corresponding marginal sample proportions $\hat{\pi}_r = \mathbf{p}_r = (p_{1+}, \dots, p_{I+})^T$, $\hat{\pi}_c = \mathbf{p}_c = (p_{+1}, \dots, p_{+J})^T$, while the MLE of $\theta^{(\phi)}$, $\hat{\theta}^{(\phi)}$, is not available in explicit form. For a given ϕ -function, model U_ϕ belongs to the family of homogeneous linear predictor (HLP) models [16]. It is straightforward to verify that it satisfies the two conditions of Definition 3 in [16]. In practice, it can be fitted using Lang's `mph` R-package. The standard U model, denoted in the sequel as U_0 , has the equivalent HLP model expression

$$L(\mathbf{m}_v) = \mathbf{C} \log(\mathbf{m}_v) = \mathbf{X}\boldsymbol{\beta} = \mathbf{1}_{(I-1)(J-1)} \theta^{(0)}, \quad (27)$$

where \mathbf{X} and $\boldsymbol{\beta}$ are the model's design matrix and parameter vector (scalar for U_0), respectively. Furthermore, $\theta^{(0)} = \log(\theta)$, \mathbf{m}_v is the $IJ \times 1$ vector of expected cell frequencies, corresponding to the $I \times J$ table \mathbf{m} expanded by rows, and \mathbf{C} is an $(I-1)(J-1) \times IJ$ design matrix so that the vector of all $\log(\text{LOR})$ s is derived, i.e., $\mathbf{C} \log(\mathbf{m}_v) = (\theta_{11}^{(0)}, \dots, \theta_{I-1, J-1}^{(0)})^T$. For more details on inference for model U_0 through HLP models, we refer to [7] (Sections 5.6 and 6.6.4). In Section 6.6.4 of [7], the approach is implemented in R for the example of Table 1 (see Section 5), while an R-function for constructing the design matrix \mathbf{C} for two-way tables of any size is provided in the web appendix of the book. This approach is easily adjusted for model U_ϕ , by replacing $\log(\mathbf{m}_v)$ in (27) by $F(\mathbf{m}_{vs})$, where \mathbf{m}_{vs} is the $IJ \times 1$ vector with entries $\frac{m_{ij}}{m_{i+}m_{+j}/n} = \frac{\pi_{ij}}{\pi_{i+}\pi_{+j}}$, expanded by rows.

Under U_ϕ , all 2×2 subtables, formed by any successive rows and any successive columns, share the same ϕ -scaled local association $\theta^{(\phi)}$. It is of practical interest to have estimators of this common local association, alternative to the MLEs, that are provided in explicit forms. One such estimator, based on the sample version of (17), is given by

$$\tilde{\theta}^{(\phi)} = \zeta(\mathbf{p}, \boldsymbol{\mu}, \boldsymbol{\nu}) \Delta_1 \Delta_2 = \Delta_1 \Delta_2 \sum_{i,j} p_{i+} p_{+j} \mu_i \nu_j F\left(\frac{p_{ij}}{p_{i+} p_{+j}}\right). \quad (28)$$

Another option is

$$\bar{\theta}^{(\phi)} = \frac{1}{(I-1)(J-1)} \sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \theta_{ij}^{(\phi)}(\mathbf{p}), \quad (29)$$

which is the mean of the sample ϕ -scaled local association measures (24). For the power-divergence based models U_λ , estimators (28) and (29) are denoted by $\tilde{\theta}^{(\lambda)}$ and $\bar{\theta}^{(\lambda)}$, respectively. For the U correlation model, derived for $\lambda = 1$ and denoted thus by U_1 , estimator (28) takes the form

$$\tilde{\theta}^{(1)} = \Delta_1 \Delta_2 \sum_{i,j} \mu_i \nu_j p_{ij} = \Delta_1 \Delta_2 r, \quad (30)$$

where r is the sample correlation between the row and column scores (compare to (19)).

Under the U_ϕ model, $\theta^{(\phi)}$ is the single association parameter, measuring the strength of the local association that is uniform across the table. Furthermore, $\theta^{(\phi)} = 0 (\Leftrightarrow \zeta = 0)$ if and only if the

independence (I) model holds. Since model I is nested in U_ϕ , the following test of independence, conditional on U_ϕ , can be considered

$$G^2(I|U_\phi) = G^2(I) - G^2(U_\phi), \quad (31)$$

which is asymptotically χ^2_1 distributed. This test is well-known for the standard U_0 model (see [17] and Section 6.3 in [7]).

Remark 3. For model U_0 , Tomizawa [18] proposed a measure of divergence from uniform association, based on the KL-divergence, taking values in the interval $[0, 1)$ and being equal to 0 if and only if U_0 holds. He constructed also asymptotic confidence interval for this measure, provided U_0 does not hold. Conde and Salicrú [19] extended his work by considering such a measure based on the ϕ -divergence and developed asymptotic inference for it, and also for the case that U_0 holds. Their approach and measures should not be confused with the approach followed here. They aim at detecting departures from U_0 (in favor of a non-uniform association structure) while here we focus on measuring the strength of uniform association, provided that U_0 (or U_ϕ) holds.

Tomizawa [18] as well as Conde and Salicrú [19] based the estimation of their measures on the following closed-form estimator of $\theta^{(0)}$ under U_0

$$\theta_*^{(0)} = \log \left(\frac{\sum_{i=1}^{I-1} \sum_{j=1}^{J-1} \theta_{ij}^{(0)}(\mathbf{p})}{(I-1)(J-1)} \right). \quad (32)$$

Remark 4. For square contingency tables with commensurable classification variables, analogous to the measure of departure from U_0 (see Remark 3), Tomizawa et al. [20] introduced a measure of departure from complete symmetry relying on the power divergence and Tomizawa [21] a measure of departure from marginal homogeneity. Kateri and Papaioannou [22] extended these measures to corresponding ϕ -divergence based measures and proposed further ϕ -divergence based measures of departure from the QS and triangular symmetry models. The work of Menéndez et al. [23,24] is also related.

5. Illustrations

We revisit a classical data set of Grizzle [25], provided in Table 2, which is adequately modelled by the U_0 model (see [18,19]). Our second data set in Table 1 corresponds to a study in [26] and provides strong evidence in favor of U_0 (see [7]). The maximum likelihood estimates of the expected cell frequencies under U_0 are also given in parentheses in Tables 1 and 2.

Table 1. Students' survey about cannabis use at the University of Ioannina, Greece (1995). The maximum likelihood estimates of the expected cell frequencies under the U_0 model are given in parentheses.

Alcohol Consumption	I Tried Cannabis...			
	Never	Once or Twice	More Often	Total
at most once/month	204 (204.4)	6 (5.7)	1 (0.9)	211
twice/month	211 (211.4)	13 (13.1)	5 (4.5)	229
twice/week	357 (352.8)	44 (48.8)	38 (37.4)	439
more often	92 (95.3)	34 (29.4)	49 (50.3)	175
Total	864	97	93	1054

Table 2. Cross-classification of duodenal ulcer patients according to operation and dumping severity. The maximum likelihood estimates of the expected cell frequencies under the U_0 model are given in parentheses.

Operation	Dumping Severity			Total
	None	Slight	Moderate	
A	61 (62.5)	28 (26.2)	7 (7.3)	96
B	68 (62.9)	23 (30.9)	13 (10.2)	104
C	58 (61.0)	40 (35.3)	12 (13.7)	110
D	53 (53.7)	38 (36.6)	16 (16.7)	107
Total	240	129	48	417

The G^2 test statistics for the U_λ models fitted on these data, for $\lambda \rightarrow 0$ and $\lambda = 1/3, 2/3, 1$ are provided in Table 3, along with corresponding $\hat{\theta}^{(\lambda)}$, $\tilde{\theta}^{(\lambda)}$ and $\bar{\theta}^{(\lambda)}$ values, i.e., the estimates of the $\theta^{(\lambda)}$ s discussed in Section 4. For U_0 , the $\theta_*^{(0)}$ estimates (32) for Tables 1 and 2 are equal to 0.2890 and 0.7972, respectively.

We observe that all considered U_λ models are of very similar (acceptable) fit for the first example while they differ enormously for the second one. For the data of Table 1, the fit of U_0 is impressive, that of $U_{1/3}$ acceptable while $U_{2/3}$ and U_1 are of very bad fit. A reverse situation appears for another data set, given in Table 4. In this case model, U_1 can be accepted while U_0 can not (see Table 5). The maximum likelihood estimates of the expected cell frequencies under U_1 are provided in Table 4 in parentheses.

Table 3. Goodness of fit of the U_λ models for the data of Tables 1 and 2 along with estimates of the common λ -scaled local association $\theta^{(\lambda)}$ under U_λ .

Example in Table 1			
λ	G^2 (p-value)	$\hat{\theta}^{(\lambda)} / \tilde{\theta}^{(\lambda)} / \bar{\theta}^{(\lambda)}$	
0	1.47 (0.917)	0.8026/0.7817/0.7814	
1/3	7.19 (0.207)	0.6857/0.6451/0.6432	
2/3	25.21 (0.000)	0.4720/0.5853/0.5974	
1	40.76 (0.000)	0.3667/0.5732/0.6096	
Example in Table 2			
λ	G^2 (p-value)	$\hat{\theta}^{(\lambda)} / \tilde{\theta}^{(\lambda)} / \bar{\theta}^{(\lambda)}$	
0	4.59 (0.468)	0.1626/0.1665/0.1612	
1/3	4.57 (0.471)	0.1619/0.1638/0.1573	
2/3	4.55 (0.473)	0.1612/0.1616/0.1541	
1	4.52 (0.477)	0.1606/0.1599/0.1515	

Table 4. Hypothetical 4×3 data table with maximum likelihood estimates of the expected cell frequencies under U_1 (in parentheses).

160 (159.5)	8 (7.8)	1 (1.4)
198 (199.4)	16 (14.7)	14 (13.9)
310 (321.8)	40 (32.9)	50 (45.4)
161 (149.1)	12 (20.4)	30 (33.8)

Table 5. Goodness of fit of the U_λ models and maximum likelihood estimates of the common λ -scaled local association $\theta^{(\lambda)}$ under U_λ for the data of Table 4.

λ	G^2 (p -Value)	$\hat{\theta}^{(\lambda)}$
0	15.99 (0.007)	0.3082
1/3	13.52 (0.019)	0.3266
2/3	10.64 (0.059)	0.3391
1	7.95 (0.159)	0.3215

Observe (in Table 3) that the closed form estimates for the λ -scaled local associations are close to the corresponding maximum likelihood estimates in case the assumed model is of adequate fit while they diverge for models of bad fit.

6. Conclusions

We revealed the role of ϕ -divergence in modelling association in two-way contingency tables and illustrated it for the special case of uniform association in ordinal contingency tables. Targeting at pointing out the potential of this modelling approach and the generated families of models, we avoided presenting technical details, properties and inferential results for these models, which can be found in the initial sources cited.

Crucial quantities are the $\theta_{ij}^{(\phi)}$ s, the ϕ -scaled measures of local association. The generalized family of ϕ -divergence based AMs enriches the modelling options in CTA, since the pattern of underlying association structure in a table may be simplified and thus described by a more parsimonious model when considering a different scale. A crucial issue, as also pointed out by one of the reviewers, is how to decide on the scale. So far, such a decision is based on trials of various alternative options. A formal approach selecting the scale is missing. In case of a parametric family, like the one based on the power divergence, the problem can be tackled by considering λ as an unknown parameter and estimating it from the data. Such an approach has been followed in the logistic regression set-up by Kateri and Agresti [15].

It is important to realize that, due to the scale difference, $\theta_{ij}^{(\phi)}$ s are not directly comparable for different ϕ -function (or λ -values in case of the power divergence). Thus, comparisons across different ϕ s (or λ s) are possible only in terms of the corresponding expected cell frequencies or a common measure of local association evaluated on them. AMs can also be considered for other types of generalized odds ratios, like, for example, the global odds ratios. The extension of such models through the ϕ -divergence and their study is the subject of a work in progress. Inference for closed form estimators of the common $\theta^{(\phi)}$ of the U_ϕ model and comparisons among them is the content of a paper under preparation.

The conditional test of independence (31) can be based not only on U_ϕ but on LL_ϕ models as well. Another 1 *df* test of independence for ordinal classification variables is the linear trend test of Mantel (see [27]). It considers the testing problem $H_0 : \rho = 0$ vs. $H_1 : \rho \neq 0$, where ρ is the correlation between ordered scores assigned to the categories of the classification variables of the table. It is thus applicable only when the underlying association exhibits a linear trend for the assigned scores. The test of Mantel uses the test statistic $M^2 = (n-1)r^2$, which is under H_0 asymptotically χ_1^2 distributed. The way this test is linked to the above-mentioned conditional tests, in view also of (19), is interesting to be investigated further.

Throughout this paper, we assumed a multinomial sampling scheme. For the models considered here, the other two classical sampling schemes for contingency tables (independent Poisson and product multinomial) are inferentially equivalent. Furthermore, for ease of presentation, we restricted here to two-way tables. The proposed models extend straightforwardly to multi-way tables. For two- or higher-dimensional tables, the subset of models that are linear in their parameters (i.e., RC-

and $RC(M)$ -type terms are excluded) belong to the family of homogeneous linear predictor (HLP) models [16] and can thus be fitted using the R-package `mph`.

Acknowledgments: The author thanks the referees for their constructive and very useful comments that improved the paper.

Conflicts of Interest: The author declares no conflict of interest.

Abbreviations

The following abbreviations are used in this manuscript:

CTA	contingency table analysis
MLE	maximum likelihood estimator
$M\phi$ Es	minimum ϕ -divergence estimators
KL	Kullback–Leibler
GOF	goodness of fit
I	independence model
AMs	association models
LL model	linear by linear association model
R model	row effect association model
C model	column effect association model
RC model	multiplicative row-column effect association model
U model	uniform association model
df	degrees of freedom
LOR	local odds ratio
CA	correspondence analysis
CI	confidence interval
QS	quasi symmetry
HLP	homogeneous linear predictor

References

1. Stigler, S. The missing early history of contingency tables. *Annales de la Faculté des Sciences de Toulouse* **2002**, *4*, 563–573.
2. Pardo, L. *Statistical Inference Based on Divergence Measures*; Chapman & Hall: New York, NY, USA, 2006.
3. Martin, N.; Pardo, N. New families of estimators and test statistics in log-linear models. *J. Multivar. Anal.* **2008**, *99*, 1590–1609.
4. Goodman, L.A. The analysis of cross-classified data having ordered and/or unordered categories: Association models, correlation models and asymmetry models for contingency tables with or without missing entries. *Ann. Stat.* **1985**, *13*, 10–69.
5. Goodman, L.A. Some useful extensions of the usual correspondence analysis and the usual log-linear models approach in the analysis of contingency tables with or without missing entries (with Discussion). *Int. Stat. Rev.* **1986**, *54*, 243–309.
6. Cressie, N.; Read, T.R.C. Multinomial Goodness-of-Fit Tests. *J. R. Stat. Soc. B* **1984**, *46*, 440–464.
7. Kateri, M. *Contingency Table Analysis: Methods and Implementation Using R*; Birkhäuser/Springer: New York, NY, USA, 2014.
8. Greenacre, M. *Correspondence Analysis in Practice*, 2nd ed.; Chapman & Hall: London, UK, 2007.
9. Gilula, Z.; Krieger, A.M.; Ritov, Y. Ordinal association in contingency tables: some interpretive aspects. *J. Am. Stat. Assoc.* **1988**, *83*, 540–545.
10. Kateri, M.; Papaioannou, T. f -divergence association models. *LRT J. Math. Stat. Sci.* **1995**, *3*, 179–203.
11. Rom, D.; Sarkar, S.K. A generalized model for the analysis of association in ordinal contingency tables. *J. Stat. Plan. Inference* **1992**, *33*, 205–212.
12. Espindiller, M.; Kateri, M. A family of association measures for 2×2 contingency tables based on the ϕ -divergence. *Stat. Methodol.* **2016**, *35*, 45–61.
13. Kateri, M.; Papaioannou, T. Asymmetry models for contingency tables. *J. Am. Stat. Assoc.* **1997**, *92*, 1124–1131.

14. Kateri, M.; Agresti, A. A class of ordinal quasi symmetry models for square contingency tables. *Stat. Probab. Lett.* **2007**, *77*, 598–603.
15. Kateri, M.; Agresti, A. A generalized regression model for a binary response. *Stat. Probab. Lett.* **2010**, *80*, 89–95.
16. Lang, J.B. Homogeneous Linear Predictor Models for Contingency Tables. *J. Am. Stat. Assoc.* **2005**, *100*, 121–134.
17. Goodman, L.A. Association models and canonical correlation in the analysis of cross-classifications having ordered categories. *J. Am. Stat. Assoc.* **1981**, *76*, 320–334.
18. Tomizawa, S. Shannon entropy type measure of departure from uniform association in cross-classification having ordered categories. *Stat. Probab. Lett.* **1991**, *11*, 547–550.
19. Conde, J.; Salicrú, M. Uniform association in contingency tables associated to Csiszár divergence. *Stat. Probab. Lett.* **1998**, *37*, 149–154.
20. Tomizawa, S.; Seo, T.; Yamamoto, H. Power-divergence-type measure of departure from symmetry for square tables that have nominal categories. *J. Appl. Stat.* **1998**, *25*, 387–398.
21. Tomizawa, S. Measures of departures from marginal homogeneity for contingency tables with nominal categories. *Statistician* **1995**, *44*, 425–439.
22. Papaioannou, T.; Kateri, M. Measures of symmetry–asymmetry for square contingency tables. In Proceedings of the 13th Conference of the Greek Statistical Institute, Florina, Greece, 3–7 May 2000; pp. 435–444.
23. Menéndez, M.I.; Pardo, J.A.; Pardo, L. Tests based on ϕ -divergences for bivariate symmetry. *Metrika* **2001**, *53*, 15–29.
24. Menéndez, M.L.; Pardo, J.A.; Pardo, L. Tests for bivariate symmetry against ordered alternatives in square contingency tables. *Aust. N. Z. J. Stat.* **2003**, *45*, 115–124.
25. Grizzle, J.E.; Starmer, C.F.; Koch, G.G. Analysis of categorical data by linear models. *Biometrics* **1969**, *25*, 489–504.
26. Marselos, M.; Boutsouris, K.; Liapi, H.; Malamas, M.; Kateri, M.; Papaioannou, T. Epidemiological aspects on the use of cannabis among university students in Greece. *Eur. Addict. Res.* **1997**, *3*, 184–191.
27. Agresti, A. *Categorical Data Analysis*, 3rd ed.; Wiley: Hoboken, NJ, USA, 2013.



© 2018 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<http://creativecommons.org/licenses/by/4.0/>).